Organization of the teicoplanin gene cluster in Actinoplanes teichomyceticus

Sosio, M.; Kloosterman, H.; Bianchi, A.; de Vreugd, P.; Dijkhuizen, L.; Donadio, S.

Published in:
Microbiology-Sgm

DOI:
10.1099/mic.0.26507-0

Publication date:
2004

Microbiology (2004), **150**, 95–102 **DOI** 10.1099/mic.0.26507-0

Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*

Margherita Sosio,[1] Harm Kloosterman,[2] Alessandra Bianchi,[1] Peter de Vreugd,[2] Lubbert Dijkhuizen[2] and Stefano Donadio[1]

[1]Vicuron Pharmaceuticals, via R. Lepetit 34, 21040 Gerenzano, Italy

[2]Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9751 NN Haren, The Netherlands

Correspondence
Margherita Sosio
msosio@vicuron.it

Received 22 May 2003
Revised 12 September 2003
Accepted 3 October 2003

The glycopeptide teicoplanin is used for the treatment of serious infections caused by Gram-positive pathogens. The *tcp* gene cluster, devoted to teicoplanin biosynthesis in the actinomycete *Actinoplanes teichomyceticus*, was isolated and characterized. From sequence analysis, the *tcp* cluster spans approximately 73 kb and includes 39 ORFs participating in teicoplanin biosynthesis, regulation, resistance and export. Of these, 34 ORFs find a match in at least one of the five glycopeptide gene clusters previously characterized. Putative roles could be assigned for most of the *tcp* genes. The two glycosyltransferases responsible for attaching amino sugars to amino acids 4 and 6 of the teicoplanin aglycon were overexpressed in *Escherichia coli* and characterized. They both recognize *N*-acetylglucosamine as the substrate. tGtfA can add a sugar residue in the presence or absence of *N*-acetylglucosamine at amino acid 4, while tGtfB can only glycosylate the teicoplanin aglycon.

INTRODUCTION

The two clinically used glycopeptide antibiotics, teicoplanin and vancomycin, interfere with bacterial cell wall synthesis by binding to the D-alanyl-D-alanine termini of peptidoglycan precursor, thus sequestering intermediate(s) during bacterial cell wall formation. Glycopeptides are often a last resort antibiotic for tough-to-treat infections by Gram-positive bacteria. However, their effectiveness is threatened by the possibility that the high-level resistance to glycopeptides frequent among enterococci may eventually become widespread in methicillin-resistant *Staphylococcus aureus*. Chemical modification of glycopeptides has led to the development of semi-synthetic glycopeptide derivatives, with improved activity, expanded antibacterial spectrum or better pharmacokinetics (Malabarba & Ciabatti, 2001).

In recent years, considerable progress has been made in understanding glycopeptide biosynthesis. Sequence information is now available for five gene clusters: chloroeremomycin (van Wageningen *et al.*, 1998), balhimycin (Pelzer *et al.*, 1999; Recktenwald *et al.*, 2002), complestatin (Chiu *et al.*, 2001), A47934 (Pootoolal *et al.*, 2002) and A40926 (Sosio *et al.*, 2003). Characterization of several gene products overproduced in *Escherichia coli* and gene inactivation experiments in the producing strain (reviewed by Hubbard & Walsh, 2003) have established the details of many biosynthetic steps. Most of this work has been carried out on the pathways of chloroeremomycin and balhimycin, two vancomycin-type glycopeptides.

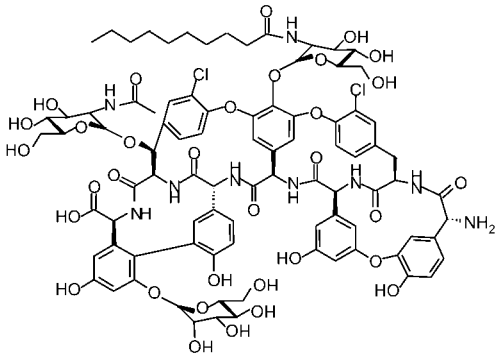
Teicoplanin (Fig. 1, compound 1), produced by the actinomycete *Actinoplanes teichomyceticus* ATCC 31131, consists of a complex of closely related molecules differing in the nature of the acyl residue. A relevant structural feature of teicoplanin is the presence of an *N*-acetylglucosamine (GlcNAc) and an *N*-acylglucosamine residue attached to amino acids 6 and 4, respectively. Glycosyltransferases (GTFs) able to attach this type of sugar to glycopeptide aglycons have not been described yet (Solenberg *et al.*, 1997; Losey *et al.*, 2001, 2002). With the objective of understanding the tailoring steps involved in decorating teicoplanin-type aglycons, we describe here the entire *tcp* cluster and report the overexpression of two active *tcp* GTF enzymes.

Abbreviations: DPG, dihydroxyphenylglycine; GTF, glucosyltransferase; HPG, *p*-hydroxyphenylglycine; βHT, β-hydroxytyrosine; NRPS, non-ribosomal peptide synthetase; TFA, trifluoroacetic acid.

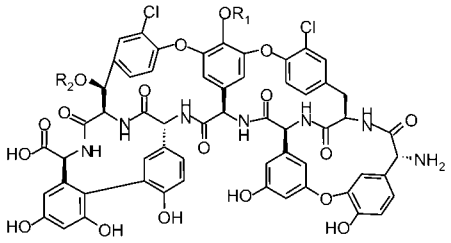
The GenBank accession number for the *tcp* gene sequence reported in this article is AJ605139.

METHODS

Isolation of the *tcp* cluster. Using total DNA of the actinomycete *Actinoplanes teichomyceticus* ATCC 31131, a cosmid library in the pWE15 vector was made by the Clontech Laboratories custom library service. The *tcp* cluster was identified as described previously (Sosio *et al.*, 2000). Cosmids were sequenced by the Seqlab custom sequencing service. The DNA sequence was analysed with programs from the Wisconsin Package (version 10, Accelrys). Each coding sequence was then compared with the *bal*, *cep*, *com*, *dbv* and *sta*



teicoplanin, **1**



compound	R_1	R_2
2	H	H
3	NAG	H
4	H	NAG
5	NAG	NAG

Fig. 1. Structures of teicoplanin and related (pseudo)-aglycons.

clusters and searched against GenBank. Most likely start codons were established by multiple alignment of related sequences, whenever possible, or by searching for ribosome-binding sites.

Expression of GTFs. The genes encoding the two *tcp* GTFs (tGtfA and tGtfB) were amplified from cosmid DNA, using the following primer pairs: 5′-GCGCATATGCGCGTGCTGTTTTCGTCC-3′ and 5′-GCAGATCTTCACGCGGGAACCGACGATC-3′ for *gtfA*, 5′-GCGCATATGCGTGTGTTGTTGTCGACGAC-3′ and 5′-CGCAGATCTCACGCGGAGACCGACGATCTCT-3′ for *gtfB*. Forward and reverse primers introduced an *Nde*I and a *Bgl*II restriction site (underlined), respectively, into the PCR fragment. PCR reactions were carried out using Vent DNA polymerase (New England Biolabs) as described by the supplier. After *Nde*I and *Bgl*II digestion, the PCR fragments were cloned into the corresponding sites of the pET3B and pET15B expression vectors (Novagen), yielding the following constructs: ptGTFA, ptGTFB, $pHis_6$tGTFA and $pHis_6$tGTFB. All constructs were sequenced to ensure that no amplification errors had been introduced. These plasmids were introduced into *E. coli* BL21(DE3)-star (Stratagene). Subsequently, transformants were transferred to LB medium supplemented with 1 M sorbitol and incubated for 24 h at 30 °C. Cultures were centrifuged, washed and resuspended in 50 mM Tris/HCl buffer (pH 9). Cells were disrupted by two passages through a French pressure cell operated at 140 MPa. Crude extracts were prepared by centrifugation for 30 min at 40 000 ***g***.

GTF activity assays and product analysis. Enzymic activity was assayed in a total volume of 100 µl, containing 1 µmol UDP-GlcNAc and 100 nmol aglycon in 50 mM Tris/HCl buffer (pH 9) and a variable amount of cell extracts containing either or both tGtfA and tGtfB proteins. Reactions were incubated at 37 °C. After reaction termination on ice, products were identified with an analytical HPLC using an AlphaBond C18 column (250 × 4·6 mm, Alltech), applying a linear gradient from 1 % (v/v) acetonitrile in 0·1 % TFA to 33 % acetonitrile in 0·1 % TFA over 15 min at a 1 ml min^{-1} flow rate, with UV detection at 280 nm. Products formed in assays with combined extracts were analysed using a linear gradient from 1 % acetonitrile in 0·1 % TFA to 12 % acetonitrile in 0·1 % TFA over 15 min. New product peaks were analysed by HPLC-MS, by delivering 10 % of the post HPLC-column flow to a Navigator quadrupole mass spectrometer (Finnigan). Mass spectral analysis was performed in the positive-ion mode of an ESI mass spectrometer, operating over a range of 1150–1650 a.m.u. Nitrogen was used as nebulizing and drying gas. Ions were accelerated in a 25 V electric field. The teicoplanin aglycon and (pseudo)aglycon (compounds 2 and 4 in Fig. 1, respectively) have been described (Malabarba *et al.*, 1984) and were kindly provided by Sonia Maffioli (Vicuron Pharmaceuticals).

RESULTS

Organization of the *tcp* cluster

From a 110 kb segment from the *A. teichomyceticus* genome (Sosio *et al.*, 2000), we sequenced a contiguous stretch of 89 713 nt. The probable boundaries of the *tcp* cluster were established from the deduced functions of the *tcp* gene products. On its left end (Fig. 2), the *tcp* cluster is likely to be delimited by ORF1, encoding a homologue of MurF, the D-alanyl-D-alanine-adding enzyme involved in cell wall formation (Table 1). On the right side, the *tcp* cluster is probably delimited by ORF39, encoding a type II thioesterase (Table 1). Consequently, a total of 39 ORFs define the *tcp* cluster, which spans almost 74 kb. The genetic organization of the *tcp* cluster is depicted in Fig. 2, and the relevant features of the 39 *tcp* ORFs are summarized in Table 1. The functions of the *tcp* genes were established by sequence comparison with homologues from other glycopeptide clusters. This analysis is reported below, grouping the *tcp* ORFs by their putative roles.

Synthesis of specialized amino acids

Apart from the Tyr residue at position 2, the heptapeptide skeleton of teicoplanin consists of non-proteinogenic amino acids: three *p*-hydroxyphenylglycine (HPG) residues (at positions 1, 4 and 5); two dihydroxyphenylglycine (DPG) at positions 3 and 7; and one *β*-hydroxytyrosine (βHT) at position 6. Seven *tcp* genes participate in the synthesis of

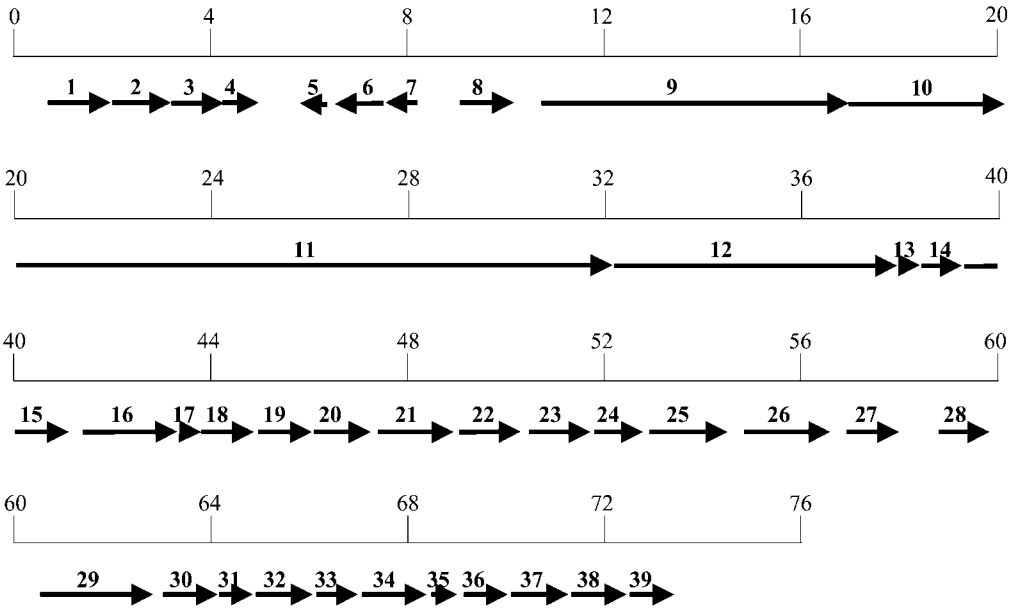


Fig. 2. Genetic organization of the *tcp* cluster. Each ORF is represented by an arrow, and numbered as in Table 1. Numbers on the scale bars indicate sequence coordinates (in kb).

HPG and DPG. ORF37 and ORF38 encode *p*-hydroxymandelate synthetase (HmaS) and *p*-hydroxymandelate oxidase (Hmo), respectively, for HPG formation (Hubbard *et al.*, 2000; Li *et al.*, 2001); ORFs 30 to 33 encode DpgA to DpgD, respectively, required for DPG synthesis (Chen *et al.*, 2001; Pfeifer *et al.*, 2001); and ORF36 encodes the Tyr-dependent aminotransferase HpgT, required for the transamination of both *p*-hydroxyphenylglyoxylate and 3,5-dihydroxyphenylglyoxylate, to yield HPG and DPG, respectively (Hubbard *et al.*, 2000; Pfeifer *et al.*, 2001). Thus, synthesis of HPG and DPG requires the same seven functions in all glycopeptide clusters that direct the synthesis of these two amino acids.

Synthesis of βHT probably involves the product of *tcp* ORF25, a putative non-haem iron dioxygenase (Table 1). Homologues of this ORF are present in the *dbv* and *sta* clusters, but not in the *bal* and *cep* clusters. Thus, a common route for βHT formation seems to operate in teicoplanin, A47934 and A40926 biosyntheses, which differs from the balhimycin and chloroeremomycin case, where tyrosine is converted into βHT through the action of three enzymes (Chen & Walsh, 2001; Puk *et al.*, 2002).

Synthesis of the heptapeptide precursor

Synthesis of the teicoplanin heptapeptide precursor is carried out by a non-ribosomal peptide synthetase (NRPS). These enzyme systems are organized in modules, with each module responsible for the recognition and incorporation of one amino acid into the oligopeptide (Marahiel, 1997). Within each module, the adenylation (A) domain is responsible for substrate recognition and activation, the thiolation (T) domain for forming a thioester bond with the cognate amino acid and elongating peptide, and the condensation (C) domain for catalysing peptide bond formation. In addition, an epimerization (E) domain is found in those modules epimerizing L-amino acids into their D-forms. The last module is usually completed by a thioesterase (Te) domain, which hydrolyses the thioester bond linking the completed peptide to the NRPS.

Four proteins, encoded by ORFs 9 to 12, constitute the seven-module teicoplanin NRPS. The genetic organization and domain composition of these ORFs indicate that ORF9, encoding the A-T-C-A-T-E domains, specifies modules 1 and 2; ORF10 (C-A-T) module 3; ORF11 (C-A-T-E-C-A-T-E-C-A-T) modules 4 to 6; and ORF12 (C-A-T-X*-Te, where X* denotes an atypical C or E domain of unknown function, so far present in all glycopeptide NRPSs) module 7. Thus, the *tcp* NRPS shows an overall domain composition identical to that of the *bal*, *cep* and *dbv* systems. Only the *com* and *sta* NRPSs present an E domain at the end of module 3, unexpected from the L-stereochemistry of the third amino acid of the heptapeptide in all glycopeptides.

Three other *tcp* ORFs are likely to participate in heptapeptide synthesis. ORFs 13 and 17 (Table 1) each encode a 69 aa polypeptide of unknown function, which is specified by an ORF located downstream of the NRPS gene encoding module 7 in the other glycopeptide clusters. The two polypeptides are completely identical over the first 60 aa, and present only 6 amino acid substitutions over their entire length. In addition, also the first 181 nt of the two ORFs are perfectly identical, suggesting that one copy arose from duplication of the other. The other ORF that may be indirectly involved in heptapeptide synthesis is ORF39,

Table 1. *tcp* ORFs

tcp ORF*		Glycopeptide clusters†					GenBank‡		Proposed function
ORF	Size§	*bal*	*cep*	*com*	*dbv*	*sta*	Match	Score	
1	447					+			D-Alanyl-D-alanine adding enzyme
2	405					+			D-Lactate dehydrogenase
3	345					+			D-Alanine : D-lactate ligase
4	202					+			D-Alanyl-D-alanine dipeptidase
5	179						ZP_00050635.1	9e−13	Unknown
6	363			+	+	+			Sensory kinase
7	226			+	+	+			Response regulator
8	393	+	+						Glycosyltransferase GtfA
9	2076	+	+	+	+	+			NRPS, modules 1–2
10	1061	+	+	+	+	+			NRPS, module 3
11	4067	+	+	+	+	+			NRPS, modules 4–6
12	1865	+	+	+	+	+			NRPS, module 7
13	69	+	+	+	+	+			Unknown
14	273	+	+		+				Unknown
15	593				+				Mannosyltransferase
16	646	+	+	+	+	+			ABC transporter
17	69	+	+	+	+	+			Unknown (related to *tcp* ORF13)
18	391	+	+	+	+	+			OxyA, cross-linking aa 2-4
19	384				+	+			Oxy, cross-linking aa 1-3
20	398	+	+	+	+	+			OxyB, cross-linking aa 4-6
21	506	+	+	+	+	+			Halogenase
22	392	+	+		+	+			OxyC, cross-linking aa 5-7
23	408	+	+		+				Glycosyltransferase GtfB
24	323				+				Unknown
25	530				+	+			β-Hydroxylase
26	598						BAB69270.1	5e−88	Acyl-CoA ligase
27	351						NP_228154.1	8e−96	Chorismate mutase
28	329	+	+	+	+	+			Regulator
29	792						AAC38065		Regulator
30	373	+	+		+	+			DpgA
31	221	+	+		+	+			DpgB
32	433	+	+		+	+			DpgC
33	268	+	+		+	+			DpgD
34	453	+	+	+	+	+			Integral membrane transporter
35	198						NP_505710.1	7e−42	GTP cyclohydrolase
36	365	+	+	+	+	+			Aminotransferase HpgT
37	351	+	+	+	+	+			HmaS
38	364	+	+	+	+	+			Hmo
39	250				+				Type II thioesterase

**tcp* ORFs as in Fig. 2.

†Homologues in other glycopeptide clusters: '+' denotes presence.

‡Best match found by BLAST searches in GenBank, other than glycopeptide clusters, with match referring to the accession number and score to the probability score of the match.

§Length in amino acid residues.

which encodes a type II thioesterase (Table 1). The proposed role for these thioesterases is to hydrolyse misprimed or misacylated T domains (Heathcote *et al.*, 2001; Schwarzer *et al.*, 2002). A type II thioesterase-encoding gene is also present towards one end of the *dbv* cluster, but not in the *bal*, *cep*, *com* or *sta* clusters.

Cross-linking and halogenation of the aromatic residues

Teicoplanin contains three ether links (between amino acids 1-3, 2-4 and 4-6) and one C–C link (between amino acids 5-7) joining its aryl groups. These cross-linking reactions are

carried out by P450 mono-oxygenases (Bischoff *et al.*, 2001a, b), and the *tcp* cluster, like the *sta* and *dbv* clusters, contains four such *oxy* genes, designated ORFs 18, 19, 20 and 22 (Table 1). It is therefore reasonable to assume that each Oxy protein carries a distinct cross-linking reaction. On the basis of the model proposed for the *bal* Oxy proteins (Bischoff *et al.*, 2001b) and using sequence similarity as the criterion, it is likely that ORF18 (77 % identical to *bal*-OxyA, but no more than 42 % identical to the other Oxy proteins) is involved in the cross-linking of the aromatic residues of amino acids 2 and 4; ORF20 (76 % identical to *bal*-OxyB, but no more than 50 % identical to the other Oxy proteins) in cross-linking of amino acids 4 and 6; and ORF22 (70 % identical to *bal*-OxyC, but no more than 47 % identical to the other Oxy proteins) for amino acids 5 and 7. By exclusion, ORF19 should be involved in cross-linking of amino acids 1-3.

Teicoplanin contains two chlorine atoms, one on the Tyr residue at position 2 and the other on the βHT residue at position 6. The product of ORF21 (Table 1) is highly related to other halogenases involved in chlorination of aromatic residues in glycopeptide biosynthesis (Puk *et al.*, 2002; Sosio *et al.*, 2003). Thus, a single halogenase should modify both residues.

Addition of sugars

Teicoplanin contains an *N*-acylglucosamine attached to the hydroxyl group of amino acid 4; a GlcNAc attached to the β-hydroxy of amino acid 6; and a mannose residue attached to one of the hydroxyl groups of amino acid 7. The roles of the two NDP-dependent GTFs encoded by ORFs 8 and 23 (Table 1) are described below. As proposed for the *dbv* cluster (Sosio *et al.*, 2003), mannosylation is likely to be carried out by a different type of GTF, unrelated to the tGtfA and tGtfB proteins described below. The putative mannosyltransferase encoded by the *tcp* cluster is specified by ORF15 (Table 1).

Resistance, export, regulation and unknown functions

The above analysis accounts for 23 of the 39 *tcp* genes. Seven of the remaining 16 ORFs are grouped at the left end of the *tcp* cluster, where the *vanHAX*-like glycopeptide-resistance cassette (ORFs 2 to 4) is located next to a *murF* homologue (Sosio *et al.*, 2000). Thus, the *sta* and *tcp* cluster share the same set of four genes at one end of the cluster, although they are differently organized. At least in the A47934 producer *Streptomyces toyocaensis*, these genes are important in conferring glycopeptide resistance (Pootoolal *et al.*, 2002). In vancomycin-resistant enterococci, homologues of ORFs 2 to 4 are involved in the synthesis of the D-alanyl-D-lactate moiety at the termini of the pentapeptide chains in nascent peptidoglycan, thus reducing the extent of glycopeptide binding to its molecular target (Evers *et al.*, 1996). ORFs 2 to 4 are therefore likely to be involved in conferring some level of teicoplanin resistance to the producing strain. It remains to be established whether ORF1, specifying a putative D-alanyl-D-alanine-adding enzyme, plays any role in conferring glycopeptide resistance.

ORFs 16 and 34 are likely to specify export functions, since they are predicted to encode an ABC-type and an ion-dependent transmembrane transporter, respectively (Table 1). ORFs 6 and 7 encode the response regulator and the sensor kinase, respectively, of a likely two-component signal transduction system (Table 1). The product of ORF28 belongs to the family of StrR-type positive transcriptional regulators, while ORF29 specifies a positive regulator of the LuxR family (Table 1).

ORFs 5, 14, 24, 26, 27 and 35 complete the list of *tcp* genes (Table 1). ORF5 is related to a protein of unknown function from *Amycolatopsis orientalis*, which however lies outside the *cep* cluster. ORFs 14 and 24 find matches in other glycopeptide clusters: the former is specified by the *bal*, *cep* and *dbv* clusters, while the latter only by the *dbv* cluster. ORF26 encodes a putative acyl-CoA ligase, while ORF27 encodes a putative 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of the AroA type, unrelated to the equivalent polypeptide encoded by the *cep* cluster. Finally, ORF35 shows considerable sequence identity to type I GTP cyclohydrolases.

Expression of GTFs

The gene composition of the *tcp* cluster makes it very likely that this region is indeed involved in the synthesis of a glycopeptide antibiotic. Since no methods for the genetic manipulation of *A. teichomyceticus* have been described, we sought to provide direct evidence for the function of selected *tcp* genes through heterologous expression. Of the two NDP-dependent GTFs identified in the *tcp* cluster, ORF8 is most related to GtfA proteins, typically glycosylating amino acid 6, and ORF23 to GtfBs, which specifically glycosylate amino acid 4. In order to establish their roles, both genes were cloned in the expression vectors pET3b and pET15B (the latter introducing an N-terminal His_6 tag). SDS-PAGE analysis of centrifuged *E. coli* extracts revealed that both proteins, without the N-terminal His-tag, were expressed in *E. coli* cultivated in LB at 37 °C, but large amounts of insoluble protein were present in the pellet fractions. Also the His-tagged version of tGtfB yielded high amounts of insoluble protein, while no His_6-tGtfA expression was observed in *E. coli*. Addition of 1 M sorbitol to the growth medium and decreasing the cultivation temperature from 37 to 30 °C resulted in a clear shift from insoluble to soluble proteins for tGtfA, tGtfB and His_6-tGtfB (data not shown). However, the altered cultivation conditions did not lead to detectable His_6-tGtfA protein. The use of compatible solutes, such as sorbitol, to stimulate correct protein folding and counteract inclusion body formation, has been described previously (Barth *et al.*, 2000).

Incubations of *E. coli* extracts expressing tGtfA with teicoplanin aglycon (Fig. 1, compound 2) and UDP-GlcNAc resulted in the appearance of a new HPLC product peak

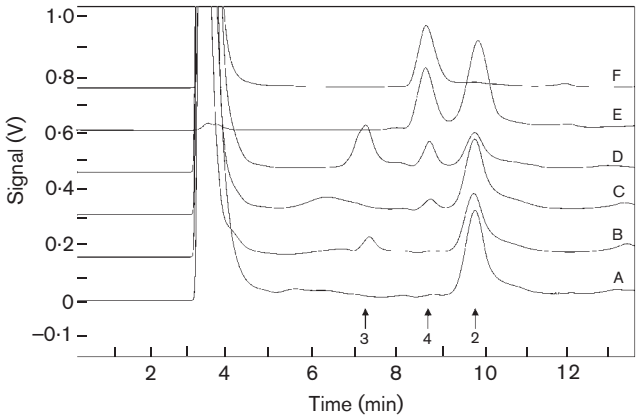


Fig. 3. GTF analysis. HPLC analysis of the incubation mixtures containing teicoplanin aglycon and UDP-GlcNAc as substrates, and either *E. coli*/pET3b (A), /ptGTFB (B), /ptGTFA (C) or combined /ptGTFB and /ptGTFA (D) extracts. Authentic compounds 4 and 2 are shown in (E). HPLC analysis of an incubation mixture containing authentic compound 4 and UDP-GlcNAc as substrates with *E. coli*-ptGTFB is shown in (F). All analyses were performed after 4 h incubation. Peaks are numbered in accordance with compound numbering of Fig. 1.

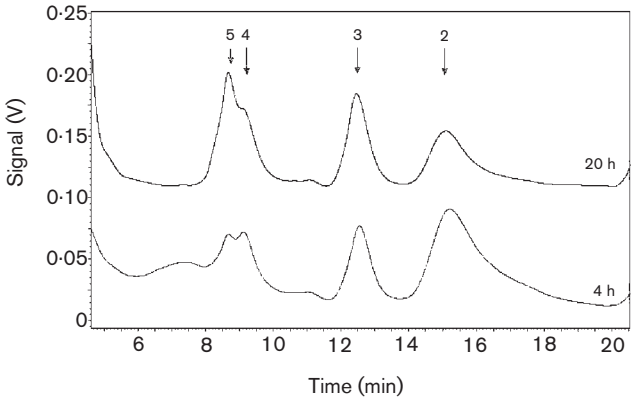


Fig. 4. Analysis of double-glycosylated aglycon. HPLC analysis (using a shallower gradient as described under Methods) of the products obtained from the combined *E. coli*/ptGTFB and /ptGTFA extracts, after 4 and 20 h incubation. Peak numbers are in accordance with compound numbering of Fig. 1.

(Fig. 3, trace C). No product was observed in incubations of extracts derived from *E. coli*-pET3b cells (Fig. 3, trace A). The retention time of the tGtfA-dependent peak coincided with that of authentic compound 4 (Fig. 3, trace E). Consistently, HPLC-MS experiments indicated that the molecular mass of the tGtfA product peak was 1400·3, confirming that tGtfA transforms compound 2 into compound 4 (Fig. 1). Incubation of the *E. coli* cell extracts expressing tGtfB or His_6-tGtfB with teicoplanin aglycon and UDP-GlcNAc revealed the appearance of a novel peak (Fig. 3, trace B), with a different retention time from compound 4 but with identical molecular mass. This result suggests that tGtfB adds a GlcNAc moiety to a different position in the teicoplanin aglycon, most likely yielding compound 3 in Fig. 1. To determine whether the combined action of tGtfA and tGtfB produces a double glycosylated aglycon, crude extracts of both proteins were combined in one assay. HPLC analysis showed that this incubation yielded three product peaks, two of which corresponded to the products observed previously in single tGtf-incubations. The retention time of the novel product peak was very close to that of compound 3 (Fig. 3, trace D). To achieve a reasonable separation of both compounds, a shallow gradient was applied, which yielded retention times of 9·2 and 8·7 min for compound 3 and the third product peak, respectively (Fig. 4). MS analysis of the 8·7 min product peak yielded a molecular mass of 1603·4, as expected from compound 5 (Fig. 1). The abundance of this peak increased with incubation time (Fig. 4).

When compound 4 replaced the teicoplanin aglycon in the incubation mixtures, tGtfB was unable to produce a new HPLC peak (Fig. 3, trace F), suggesting that compound 4 is not a substrate for this enzyme. As expected, no conversion of compound 4 was observed with tGtfA (data not shown).

When cell extracts containing either tGtfA or tGtfB were incubated with vancomycin aglycon and UDP-GlcNAc, no HPLC product peak was observed, not even after 24 h incubation. This is in contrast with the *cep*-encoded GtfB enzyme, which also glycosylates the teicoplanin aglycone at amino acid 4 (Losey *et al.*, 2001, 2002). UDP-sugar substrate specificity was tested for both tGtf enzymes in incubations with teicoplanin aglycon. The activated sugars tested were: UDP-*N*-acetylgalactosamine, U/TDP-glucose and UDP-galactose. No conversion of teicoplanin aglycon was found with any of these activated sugars.

DISCUSSION

We have described here the complete *tcp* cluster and provided a first characterization of the activities and substrate specificities of two NDP-sugar-dependent GTFs encoded by this cluster. A total of five gene clusters (*bal*, *cep*, *dbv*, *sta* and *tcp*) involved in the synthesis of antibacterial glycopeptides have now been described, in addition to the *com* cluster directing formation of an anti-complement glycopeptide. These clusters derive from four different genera of actinomycetes (*Actinoplanes*, *Amycolatopsis*, *Nonomuraea* and *Streptomyces*). It should be noted that, despite the very similar chemical structures of the two glycopeptides A40926 and teicoplanin, only 28 of the 39 ORFs specified by the *tcp* cluster find a homologue in the *dbv* cluster (Table 1).

One unique feature of the *tcp* cluster is the lack of an ORF encoding a putative prephenate dehydrogenase (PDH). PDH-encoding ORFs have been observed so far in the five other glycopeptide clusters, as well as in the *cda* (Hojati *et al.*, 2002) and ramoplanin (Zazopoulos *et al.*, 2002) clusters

from *Streptomyces coelicolor* and *Actinoplanes* sp. ATCC 33076, respectively. The corresponding compounds contain an HPG residue. The presumed role of PDH is to prime HPG synthesis by generating *p*-hydroxyphenylpyruvate (HPP), the first substrate for the enzyme Hmo. It should be noted that, using Tyr as an amino donor, HpgT generates one molecule of HPP. In the actinomycetes where this has been investigated, Tyr biosynthesis requires an arogenase, which converts prephenate into arogenate (Hodgson, 2000). In addition, the *S. coelicolor* genome (Bentley *et al.*, 2002) contains a single PDH-encoding gene within the *cda* cluster. We do not know whether *A. teichomyceticus* possesses a PDH-encoding gene somewhere else in its genome. However, it should be noted that, in those glycopeptides where both HPG and DPG are utilized, HPP can be formed by transamination of Tyr during DPG formation. Thus, a PDH activity may be strictly necessary only in those cases when HPG synthesis is unlinked to that of DPG.

The gene(s) responsible for *N*-acylation of the glucosamine residue have not been identified. Contrary to our expectations, the *tcp* cluster specifies no homologue of *dbv* ORF23, which contains motifs typical of the family 3 of acyltransferases (Sosio *et al.*, 2003) and could have been a likely candidate for carrying out this reaction. Instead, the *tcp* cluster specifies a putative acyl-CoA ligase, encoded by ORF26 (Table 1). If *N*-acylation is carried out by homologous genes during teicoplanin and A40926 biosynthesis, then the *dbv* and *tcp* clusters should share a gene absent from the other glycopeptide clusters. The only ORFs of unassigned function fitting this requirement are represented by *tcp* ORF24 and *dbv* ORF8. However, these sequences show no homology in database searches. An alternative possibility is that *N*-acylation is carried out through different mechanisms in A40926 and teicoplanin formation. The existence of two different pathways for βHT formation represents a precedent in glycopeptides.

We observed that compound 5 is formed by the combined action of tGtfA and tGtfB while compound 4 is not a substrate for tGtfB. This suggests that tGtfA can glycosylate the teicoplanin molecule independently of the glycosylation at amino acid 4, while tGtfB can act only on the aglycon. This would imply that tGtfB acts before tGtfA during teicoplanin biosynthesis in *A. teichomyceticus*. *In vivo* studies with the balhimycin producer indicated that a *bgtfB* deletion mutant produced only the balhimycin aglycon (Pelzer *et al.*, 1999), suggesting that bGtfA is unable to attach epivancosamine to amino acid 6. If this is true also *in vitro*, it would mean that bGtfA, unlike tGtfA, can only recognize a balhimycin pseudoaglycon glucosylated at amino acid 4.

A second finding is that tGtfB recognizes UDP-GlcNAc as the substrate. Since we did not have access to UDP-glucosamine, we do not know whether this reflects a relaxed specificity by tGtfB. The alternative possibility is that UDP-GlcNAc is the true substrate for tGtfB, and then a transacylation reaction replaces the acetyl group on this sugar with the fatty acid chain found in teicoplanin. This latter hypothesis is also consistent with the utilization of UDP-GlcNAc by several cellular pathways, while we are unaware of instances where bacterial metabolism uses an NDP-activated glucosamine.

ACKNOWLEDGEMENTS

We are grateful to Wolfgang Wohlleben and John Robinson for sharing unpublished information, and valuable discussions. We also thank G. B. Panzone for providing the teicoplanin aglycone, and G. J. Euverink for MS analysis. This work was partially supported by a grant from the EU (QLK3-1999-00650) and by the Italian CNR (PF Biotecnologie).

REFERENCES

Barth, S., Huhn, M., Matthey, B., Klimka, A., Galinski, E. A. & Engert, A. (2000). Compatible-solute-supported periplasmic expression of functional recombinant proteins under stress conditions. *Appl Environ Microbiol* **66**, 1572–1579.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M. & 40 other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.

Bischoff, D., Pelzer, S., Höltzel, A., Nicholson, G., Stockert, S., Wohlleben, W., Jung, G. & Süßmuth, R. (2001a). The biosynthesis of vancomycin-type glycopeptide antibiotics – new insights into the cyclization steps. *Angew Chem Int Ed* **40**, 1693–1696.

Bischoff, D., Pelzer, S., Bister, B., Nicholson, G. J., Stockert, S., Schirle, M., Wohlleben, W., Jung, G. & Süßmuth, R. D. (2001b). The biosynthesis of vancomycin-type glycopeptide antibiotics – the order of cyclization steps. *Angew Chem Int Ed* **40**, 4688–4691.

Chen, H. & Walsh, C. T. (2001). Coumarin formation in novobiocin biosynthesis: β-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem Biol* **74**, 1–12.

Chen, H., Tseng, C. C., Hubbard, B. K. & Walsh, C. T. (2001). Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **98**, 14901–14906.

Chiu, H. T., Hubbard, B. K., Shah, A. N., Eide, J., Fredenburg, R. A., Walsh, C. T. & Khosla, C. (2001). Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci U S A* **98**, 8548–8553.

Evers, S., Quintiliani, R., Jr & Courvalin, P. (1996). Genetics of glycopeptide resistance in enterococci. *Microb Drug Resist* **2**, 219–223.

Heathcote, M. L., Staunton, J. & Leadlay, P. F. (2001). Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem Biol* **8**, 207–220.

Hodgson, D. A. (2000). Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol* **42**, 47–238.

Hojati, Z., Milne, C., Harvey, B. & 9 other authors (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol* **9**, 1175–1187.

Hubbard, B. K. & Walsh, C. T. (2003). Vancomycin assembly: nature's way. *Angew Chem Int Ed* **42**, 731–765.

Hubbard, B. K., Thomas, M. G. & Walsh, C. T. (2000). Biosynthesis of L-*p*-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **7**, 931–942.

Li, T. L., Choroba, O. W., Charles, E. H., Sandercock, A. M., Williams, D. H. & Spencer, J. B. (2001). Characterisation of a hydroxymandelate oxidase involved in the biosynthesis of two unusual amino acids occurring in the vancomycin group of antibiotics. *Chem Commun* **2001**, 1752–1753.

Losey, H. C., Peczuh, M. W., Chen, Z., Eggert, U. S., Dong, S. D., Pelczer, I., Kahne, D. & Walsh, C. T. (2001). Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. *Biochemistry* **40**, 4745–4755.

Losey, H. C., Jiang, J., Biggins, J. B., Oberthür, M., Ye, X.-Y., Dong, S. D., Kahne, D., Thorson, J. S. & Walsh, C. T. (2002). Incorporation of glucose analogs by GtfE and GtfD from the vancomycin biosynthetic pathway to generate variant glycopeptides. *Chem Biol* **9**, 1305–1314.

Malabarba, A. & Ciabatti, R. (2001). Glycopeptide derivatives. *Curr Med Chem* **8**, 1759–1773.

Malabarba, A., Strazzolini, P., Depaoli, A., Landi, M., Berti, M. & Cavalleri, B. (1984). Teicoplanin, antibiotics from *Actinoplanes teichomyceticus* nov. sp. VI. Chemical degradation: physico-chemical and biological properties of acid hydrolysis products. *J Antibiot* **37**, 988–999.

Marahiel, M. A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chem Biol* **4**, 561–567.

Pelzer, S., Süßmuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G. & Wohlleben, W. (1999). Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob Agents Chemother* **43**, 1565–1573.

Pfeifer, V., Nicholson, G. J., Ries, J., Recktenwald, J., Schefer, A. B., Shawky, R., Schröder, J., Wohlleben, W. & Pelzer, S. (2001). A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *J Biol Chem* **276**, 38370–38377.

Pootoolal, J., Thomas, M. G., Marshall, C. G., Neu, J. M., Hubbard, B. K., Walsh, C. T. & Wright, G. D. (2002). Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc Natl Acad Sci U S A* **99**, 8962–8967.

Puk, O., Huber, P., Bischoff, D., Recktenwald, J., Jung, G., Süßmuth, R. D., Van Pee, K.-H., Wohlleben, W. & Pelzer, S. (2002). Glycopeptide biosynthesis in *Amycolatopsis mediterranei*: function of a halogenase and a haloperoxidase/perhydrolase. *Chem Biol* **9**, 225–235.

Recktenwald, J., Shawky, R., Puk, O., Pfennig, F., Keller, U., Wohlleben, W. & Pelzer, S. (2002). Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology* **148**, 1105–1118.

Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083–14088.

Solenberg, P. J., Matsushima, P., Stack, D. R., Wilkie, S. C., Thompson, R. C. & Baltz, R. H. (1997). Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*. *Chem Biol* **4**, 195–202.

Sosio, M., Bianchi, A., Bossi, E. & Donadio, S. (2000). Teicoplanin biosynthesis genes in *Actinoplanes teichomyceticus*. *Antonie Van Leeuwenhoek* **78**, 379–384.

Sosio, M., Stinchi, S., Beltrametti, B., Lazzarini, A. & Donadio, S. (2003). The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. *Chem Biol* **10**, 541–549.

van Wageningen, A. M., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., Jones, M., Jones, S. J. & Solenberg, P. J. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol* **5**, 155–162.

Zazopoulos, E., Farnet, C. M. & Staffa, A. (2002). Gene cluster for ramoplanin biosynthesis. Patent WO 0231155-A 1 18-APR-2002; Ecopia Biosciences Inc. (CA). Accession AX417445.